THE ONTARIO TEST OF ENGLISH AS A SECOND LANGUAGE (OTESL)

A Report on the Research

M. WESCHE
M. CANALE
E. CRAY
S. JONES

D. MENDELSOHN
M. TUMPANE
M. TYACKE

This research project was funded under contract by the Ministry of Colleges and Universities and the Ministry of Education, Ontario.
It reflects the views of the authors and not necessarily those of the Ministries.

Gregory Sorbara, Minister of Colleges and Universities
Alan K. Adlington, Deputy Minister of Colleges and Universities

Sean Conway, Minister of Education
Bernard J. Shapiro, Deputy Minister of Education

Ministry of Colleges and Universities
Ministry of Education

87-05507/

THE ONTARIO TEST OF ENGLISH AS A SECOND LANGUAGE (OTESL)

A Report on the Research

M. WESCHE
M. CANALE
E. CRAY
S. JONES
D. MENDELSOHN
M. TUMPANE
M. TYACKE

This research project was funded under contract by the Ministry of Colleges and Universities and the Ministry of Education, Ontario.

It reflects the views of the authors and not necessarily those of the Ministries.

Gregory Sorbara,
Minister of Colleges and Universities
Alan K. Adlington,
Deputy Minister of Colleges and Universities

Sean Conway,
Minister of Education
Bernard J. Shapiro,
Deputy Minister of Education

© QUEEN'S PRINTER FOR ONTARIO, 1987

Order information:

Publications Sales
The Ontario Institute for
Studies in Education
252 Bloor Street West
Toronto, Ontario
M5S 1V6
(416) 926-4707

Will invoice on orders over $30.00.
Other orders must be accompanied by a
cheque or money order payable to
O.I.S.E.

MGS Publications Services
880 Bay Street, 5th Floor
Toronto, Ontario
M7A 1N8
(416) 965-6015
(Toll Free) 1-800-268-7540
(Toll Free from area code 807)
Ask operator for Zenith 67200.

Order must be accompanied by a cheque
or money order payable to the
Treasurer of Ontario.

Contract 0661
ONO 3527

Canadian Cataloguing in Publication Data

Main entry under title:
The Ontario test of English as a second language (OTESL)

Co-published by: Ontario Ministry of Education.
ISBN 0-7729-2613-1

1. English language — Examinations. 2. English
language — Study and teaching as a second language
(Higher) — Ontario*. I. Wesche, Marjorie Bingham,
1941- . II. Ontario. Ministry of Colleges and
Universities. III. Ontario. Ministry of Education.

LB2367.C36057 1987 428'.0076 C87-099629-0

Administrative Coordinator D. Ready

Project Liaison E. Cummings

Administrative Assistant L. Malo

Technical Assistant H. Zuckernick

Materials Developers

G. Arbach
P. Currie
J. Doherty
A. Gray
J. Guravich
M. Henkelman
M. Hynes
L. Lebrande
M. Lee
D. Louson
C. McCarthy
M. McNerney
M. Pierce
T. Pychyl
B. Ronaghan
B. Smith
J. St. John
E. Taborek
J. Thompson

Consultants

M. Boss
B.J. Carroll
A. Davies
C. Weir

Advisory Committee

P. Allen
L. Bebout
L. Buchanan
M. Callahan
P. Evans
J. Handscombe
R. Harrison
P. Hauptman
G. Irons
J. Jewinski
J. Jones
I. Martin
S. Nishisato
D. Palmer
(CONDAT, Chairman 1984-6)
C. Reed
J. Stevens
(CONDAT, Chairman 1982-4)
H.H. Stern
M. Swain
P. Watts
J. Yalden

Field-testers

S. Beare
L. Buchanan
L. Courtney
R. Diebel
S. Firth
C. Hallitt
L. Irving
J. Jewinski
J. Jones
I. Martin
N. Naiman
S. Paribakht
J. Sims

Participating Institutions

Algonquin College
Carleton University
Eastern High School of Commerce, Toronto
George Brown College
Laurentian University
Mohawk College
Ontario Institute for Studies in Education
Ryerson Polytechnical Institute
Sir Sanford Fleming High School
University of Ottawa
University of Toronto
University of Waterloo
York University

ACKNOWLEDGEMENTS

Development of the OTESL was funded by the Ontario Ministry of Colleges and Universities under contract with the Council of Ontario Universities' Committee on Diagnostic Achievement Testing (CONDAT). THe project was administered by the Centre for Second Language Learning, University of Ottawa.

The Council of Ontario Universities
Le Conseil des Universités de l'Ontario
130 St. George Street
Suite 8039
Toronto, Ontario
M5S 2T4

CONTACT PERSONS

While development of the OTESL was a team effort involving many ESL educators, the following individuals have agreed to serve as contact persons for different aspects of the project.

Institutional Needs Analysis	E. Cray
Placement Test	M. Tumpane, E. Cray
Post Admission Test:	
General Academic (Reading)	M. Wesche
Social Sciences, Forms A,B	E. Cray, D. Ready
Science & Technology, Forms A,B	S. Jones, M. Wesche
Writing	S. Jones, E. Cray
Oral Interaction	D. Mendelsohn, M. Tumpane
Training of administrators and raters	M Canale
Information on test dissemination	M. Tyacke

E. Cray, S. Jones
Department of Linguistics
Carleton University
Colonel By Drive
Ottawa, Ontario
K1S 5B1
(613) 564-5573

D. Mendelsohn, M. Tyacke
Continuing Education
University of Toronto
158 St. George Street
Toronto, Ontario
M5S 2V8
(416) 978-6529

D. Ready, M. Wesche
Centre for Second Language Learning
University of Ottawa
Ottawa, Ontario
K1N 6N5
(613) 788-2973

M. Tumpane
English Department
George Brown College
507 College Street
Toronto, Ontario
M6G 1A8
(416) 967-1212

M. Canale
Franco-Ontarian Centre
OISE
252 Bloor Street West
Toronto, Ontario
M5S 1V6
(416) 923-6641

ABSTRACT

The Ontario Test of English as a Second Language (OTESL) is a battery of instruments and procedures for evaluating the English proficiency of second language speakers in Ontario post-secondary institutions. It was developed under contract between the Ontario Ministry of Colleges and Universities and the Council of Ontario Universities by a team of researchers from five Ontario post-secondary institutions.

The instruments serve two major purposes. One is to place students in intensive pre-admission ESL programmes, using a 45-minute **Placement Test** with listening, reading, writing and speaking components. The second is to evaluate the language skills of students already admitted to academic programmes who may still need further English instruction. This is done using a 2 1/2 hour **Post-Admission Test** which measures performance in reading, listening and writing for academic purposes and an **Oral Interaction Test.** Both instruments measure global proficiency levels according to functional descriptions on an assessment scale and also give diagnostic feedback for students and ESL programmes.

The battery was developed to specifications which were based on information obtained from a survey of the language needs of ESL students in Ontario post-secondary institutions and of institutional ESL testing needs. In response to the somewhat different language needs of science and non-science students, 2 alternate versions of the **Post-Admission Test** and **Oral Interaction** tasks were developed. Both versions of the written test include a common general academic timed-reading component and a discipline-related module (in Social Science or Science and Technology) which consists of two readings, a brief lecture on a related theme and a writing task based on information from the readings and lecture. A second form of this discipline-related component was developed for both versions. A fifteen minute **Oral Interaction Test** of English use for academic purposes is also part of the battery, and can be used independently of the written test. It is based on tasks of increasing difficulty, moving from basic social interaction to academic language use, and involves 2 trained administrators.

Developed from authentic source material the OTESL instruments reflect academic texts, discourse types and tasks which represent what ESL students in academic situations need to be able to do in English to succeed in their studies. This use of a performance approach -- in which examinees must demonstrate their L_2 knowledge in the context of tasks and situations which are similar to those for which they need the language, has been very positively received by ESL teachers, students and admission personnel involved in the project during the development and field-testing of materials. Concurrent validation measures also indicate that OTESL results correspond highly with the judgements of student proficiency made by language teachers.

TABLE OF CONTENTS

Section

I. Background and Objectives

II. Instrument Design: Format and Specifications

III. Process of Test Development

IV. Statistical Characteristics of OTESL

V. Test Distribution: Use, Maintenance and Revision: A Proposal

References

Appendices

A Needs Assessment Report

B OTESL **Post-Admission Test** Specifications

C General Assessment Scale (Carroll)
Oral Interaction Assessment Bands

D Diagnostic Feedback Sheets: Post-Admission Test Score Report
Writing Evaluation Report Form,
Oral Interaction Diagnostic Comments

E. Statistical Tables

The **OTESL User's Manual** containing information on test administration, scoring and interpretation, is bound separately, as are the test instruments. These confidential documents will be available to authorized user institutions.

I. BACKGROUND AND OBJECTIVES

The Ontario Test of English as a Second Language (OTESL) project developed out of a request by the Committee on Diagnostic Achievement Testing (CONDAT) of the Council of Ontario Universities (COU) for the development of an evaluation procedure for speakers of English as a second language (ESL) following admission to Ontario post-secondary institutions. A standardized test battery was needed by institutions to determine appropriate placement in ESL courses, to determine whether these students had adequate English proficiency to pursue their academic studies in English, and for diagnostic purposes in cases where students required further training. A common instrument across the province should also aid in student mobility among institutions and provide valuable feedback to secondary school English programmes for ESL speakers.

Several years of preliminary work by the principal investigators, other ESL educators in Ontario and CONDAT members led to a proposal for which Ministry of Colleges and Universities funding was approved in the Spring of 1983. During this period, upon Ministry request, the project was expanded to take into account the ESL testing requirements and academic programmes in the province's colleges as well as its universities.

The project officially began in September 1983, with a team of six principal investigators (subsequently enlarged to seven) representing five Ontario post-secondary institutions. It was administered at the University of Ottawa, Centre for Second Language Learning. An Advisory Committee was formalized, including many language educators who had been involved in the project's conception. Members were drawn from various institutions across Ontario and provided a broad representation of ESL educational concerns for grades 12-13, colleges and universities. An ongoing consultative process with this Committee, and with ESL professionals in

the field and their students provided the project team with an overview of the diverse needs of Ontario ESL programmes and clienteles. The OTESL institutional needs analysis, test specifications, test format, topics, procedures and tasks reflect this process.

The overall objectives of the Ontario Test of English as a Second Language are summarized below.

The test should provide valid, reliable, feasible evaluation procedures of English language proficiency for Ontario post-secondary institutions, which can be administered locally, and which can serve the following purposes:

a) determine whether students have the requisite academic English skills to be able to follow given courses of study in these institutions;
b) identify those areas of English in which students require further tuition, and to provide diagnostic feedback for students and programmes;
c) provide information for the appropriate placement of students in ESL courses and programmes;
d) provide information interpretable across institutions for such purposes as facilitation of student transfers and feedback to secondary school programmes.

II. INSTRUMENT DESIGN: FORMAT AND SPECIFICATIONS

The OTESL team first carried out a needs survey in Ontario post-secondary institutions (Appendix A) with administrators and ESL educators.

Design specifications for the **Post-Admission Test** were drawn up (Appendix B) based on the survey as well as on the following three supplementary sources:

a) research literature in the field of second language testing, particularly the testing of academic English skills,

b) a consideration of the feasibility of various testing approaches re potential time limitations for test administration and restrictions on availability of trained personnel to administer and score the test in the user institutions,

c) the combined expertise and experience of the Project Team, Advisory Committee members and our consultants in test development, administration and interpretation in situations similar to those for which OTESL is destined.

A major concern of the OTESL team has been to develop instruments which are task-valid with respect to the projected second language needs of examinees. Thus a performance testing approach has been followed, based on an analysis of the language situations and the kinds of written and oral discourse students will encounter in a Canadian academic setting, and on related specification of the language processing skills and functions which will be most essential to these students in their academic work. As has been demonstrated by the world-wide experience of

the British Council's English Language Testing Service (British Council 1982; Carroll 1981; Seaton 1985) and the Associated Examining Board (Emmett 1985; Weir 1983), direct performance tests which focus on contextualized language use rather than on knowledge of correct linguistic forms can provide an accurate, reliable and detailed picture of how prepared an ESL student is to enter a given post-secondary program in an English-language institution. In addition to the diagnostic usefulness of such tests for examinees and ESL programmes, they are particularly valuable in their overall "washback" effect on ESL programmes because they send the message to students and teachers alike that the emphasis should be on what one can do with the second language rather than what one knows about it or one's strategies for successful test-taking.

It was decided that listening, reading, speaking and writing skills would be tested in an integrated fashion through tasks which simulate situations of academic language use, such as listening to a lecture and taking notes, or synthesizing ideas from several readings and lecture material in a structured essay.

The **Post-Admission Test** is aimed at relatively advanced English proficiency levels, including as its focal range those students who possess the requisite English skills to carry on regular full-time or part-time academic studies and those who are close to this level.

It was decided to report scores on this test according to descriptive performance bands, using B. J. Carroll's General Assessment Scale (1980) as a guideline. Modified bands were subsequently developed for the **Oral Interaction Test.** (Both are found in Appendix C.)

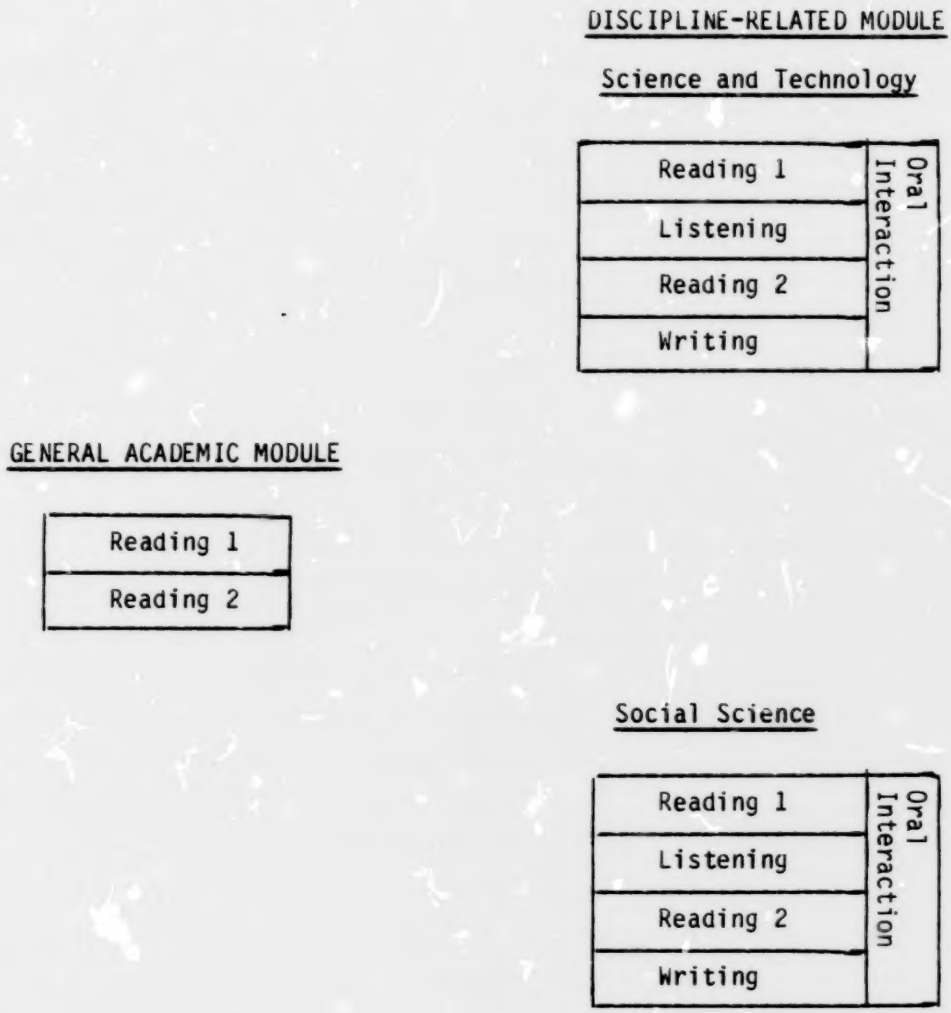
It soon became clear that the **Post-Admission Test** could not meet the needs of institutional ESL programmes for an easy-to-administer pre-admission placement instrument to sort students at lower proficiency levels, a widespread need in intensive ESL courses in both colleges and universities. Thus it was decided to develop a **Placement Test** as part of the UTESL battery, which could be used alone or with certain other components of the **Post-Admission Test** as required by different programmes. Both instruments are described in detail below.

POST ADMISSION TEST

The **OTESL Post-Admission Test** (Figure 1) is intended to provide information on the performance of academic tasks in English by ESL students who have already been admitted to an academic programme at a post-secondary institution. It exists in two forms, A and B. The test requires approximately three hours to administer and is comprised of three independent components: the General Academic component to be taken by all students (1/2 hour); a discipline-related component - the Social Science Module or the Science and Technology Module - that students may select according to their interests (2 hours); and the **Oral Interaction Test** that may be administered with alternative discipline-related stimulus material for either social science or science students (15 minutes). Both discipline-related modules have the same structure: authentic academic texts and tasks on a given theme are presented in a Reading section, then comes a Listening section, a second Reading section, and finally, a Writing section. The General Academic module uses general interest texts and charts and is composed of two Reading subsections. All tests and tasks are intended to represent levels 4 through 7 on the General Assessment Scale and its adaptation for the **Oral Interaction Test** (Appendix C). The **Oral Interaction Test** is a semi-directed interview that is adapted to the performance level of the student as it is administered. It includes four tasks of increasing difficulty: social interaction, comparison (on a familiar topic), rephrasing of factual written material (non-science or science focus) and an opinion or recommendation task.

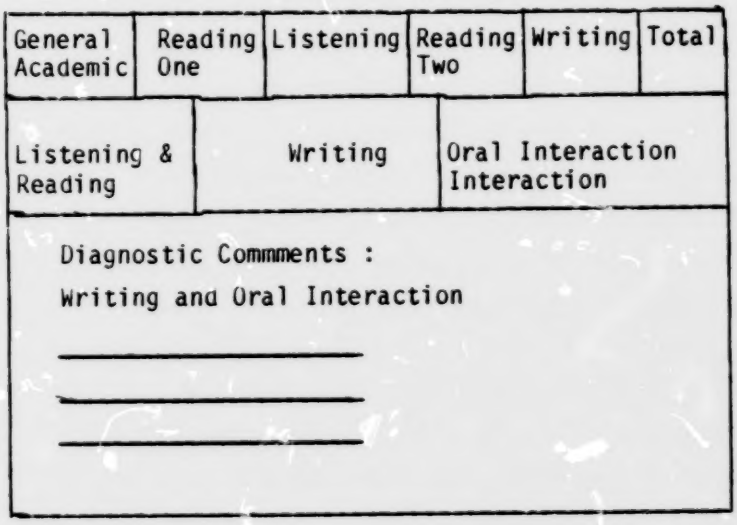
The overall structure of the **Post-Admission Test** is shown in Figure 1.

Figure 1. POST ADMISSION TEST DESIGN



A variety of item types and procedures is used in the **Post-Admission Test.** A short answer format is used where it is considered to be task valid. In other cases (e.g. Writing task, **Oral Interaction Test**), freer-form tasks are essential and require scoring by trained native speakers. The three-level score reporting procedure provides information at different levels of detail to students, teachers, and administrators. (Figure 2.)

Figure 2. SCORE REPORTING SCHEMA



Raw Scores

Profile of band scores: Listening and Reading, Writing and Oral Interaction

Diagnostic feedback sheets; Writing and Oral Interaction

The profile scores, of use to administrators, provide a summative statement about the candidates' overall ESL proficiency. They are criterion - referenced according to the proficiency bands (Appendix C). The profile scores will also be of interest to academic departments to which candidates are applying as well as to ESL programme designers. A more detailed diagnosis of micro-skills and functions can also be reported for Writing and Oral Interaction, primarily for use by ESL teachers and the students themselves, together with the raw scores for each test section. (These are found in Appendix D.)

III. PROCESS OF TEST DEVELOPMENT

The process of OTESL instrument development is briefly sketched below.

The members of the research team first developed a questionnaire which was used to gather information on the student population of speakers of English as a second language in Ontario universities and colleges.

In a four-step process, information was gathered using this instrument. First, the Admissions Offices of all Ontario universities and colleges were contacted and asked to provide information on admission criteria for non-native speakers of English: specifically, which groups had to prove English language proficiency and what tests were used to determine adequate proficiency. Admissions officers and/or ESL programme administrators were then contacted by mail or in person to gather additional information on programmes available to ESL students and on local testing procedures. Third, information was collected on the ESL student population in each programme, including countries of origin and the academic disciplines most commonly entered. Finally, administrators were asked what their testing needs were, and local tests were collected for examination by the research team.

Contact was also made with the principal investigators for the Ontario Test of English Achievement (OTEA) and the College Diagnostic Language Test (CDLT) which are post-secondary English tests for native speakers of English. Both tests were examined and discussions on test design and development were held with the members of the two test development groups.

At the same time both collection of other relevant first and second language proficiency tests and a survey of current literature on test development were begun, with particular emphasis on second language proficiency tests being developed by the British Council, British universities and independent examining boards.

The research team also investigated the ESL student population in grades 12 and 13 with the help of Advisory Committee members who work with this group. The testing procedures at the Ottawa Board of Education and the North York Board of Education were surveyed and administrators from these boards, both of which have large ESL populations, were interviewed about their testing needs.

After it had been determined that the disciplines most commonly entered by ESL students are business, economics, engineering and science, and computer science, members of the research team began to look at first-year university tests and examinations to determine what students are asked to do in introductory courses. This was done both to determine the language proficiency needs of ESL students and to locate source material for OTESL. In addition, members of the team viewed many hours of video-taped university lectures and discussed course content assignments and tests with a number of university lecturers.

In November, 1983, the research team met to determine test design and format. During this three-day meeting, the group used the information gathered during the first two months of the project to design a test format and write a set of specifications for the **Post-Admission Test.**

The decision was made to use a performance-based (work sample) approach, and to develop both a general academic and two alternate discipline-related components for students in social science versus those in science and technology related diciplines. The decision to develop two alternative components was based on work by British researchers suggesting that scores on performance tests have greater validity if students are tested on content which is related to their academic disciplines (Alderson and Urquhart 1985). A further justification was the greater utility of

domain specific tests for diagnostic purposes. The team also decided to use Carroll's language proficiency bands or adaptations of them as the basis for the scoring system of the test. This system has been used by the British Council and has been found to be appropriate for administrative and diagnostic purposes in the evaluation of academic ESL skills.

The next step of the project was to find and adapt texts, and to develop trial items. Item writers were selected and trained and then given texts prepared by the research team to develop items and instructions. During January and February, 1984, materials for the discipline-related components were developed and initial pre-testing was carried out at several universities. Following this, the research team revised and edited items and then prepared first drafts of the components. Materials for a general academic component were also developed but were found to be too easy when field-tested. The revised discipline-related components were tested with approximately 150 students at this point.

Students taking the test were asked to provide information on their native language, number of years in Canada, estimated language proficiency, and to give their impressions of the test. Sixty-five students who planned to enter Ontario universities or colleges were interviewed in depth and asked to comment on specific items and test instructions. Teachers were asked to rate the language proficiency of the students taking the test using the Carroll language proficiency bands, and to rank the students. They were also asked to provide feedback on the pilot test. The purpose of this pre-testing and data collection was to examine the validity, reliability and feasibility of the item types, the instructions and of the draft components of OTESL.

Following this pre-testing, a pilot form of the OTESL **Post-Admission Test** was completed and members of the research team wrote draft scoring and administration instructions for each component. An outline for the **Oral Interaction Test** was also completed. The draft test materials were sent to members of the Advisory Committee who were asked to comment on the structure and content of the test. Further revisions were made following the responses of members of the Advisory Committee in the spring of 1984.

Following Ministry approval of continued funding for OTESL development, the project was resumed in November 1984 after a five-month shut-down. Further revisions were made to all test materials, administration and scoring procedures and documents for the collection of validity data. The pilot test components were then field-tested at four different universities and two colleges in the spring of 1985. Statistical analyses and feedback from the field-testing led to further revisions.

During the summer of 1985, intensive work was done on the **Oral Interaction Test**, a **Placement Test** for pre-admission intensive programmes, and on the General Academic component of the **Post-Admission Test.** Detailed procedures were developed for scoring the Writing component of the **Post-Admission Test,** and workshops were held to train raters for the evaluation of Writing field-test results. Text material was selected and developed for the Social Science and Science and Technology Form B of the **Post-Admission Test.**

Ongoing field-testing and revisions of all test components continued through the fall of 1985. Once satisfactory final forms of the **Placement Test,** the **Oral Interaction Test** and Form A of the **Post-Admission** test were achieved, student samples were identified for final field-testing to gather data for the **OTESL User's Manual.** This testing was carried out from November 1985 through January 1986 in a number of secondary schools, colleges and universities. Pilot versions of the Form B discipline-related components of the **Post-Admission Test** were also tested during this time, and subsequently revised. Training procedures and a training video-tape for the **Oral Interaction Test** were finalized and tried out in the field.

During January and February 1986, the following work was completed.

Form B of the **Post-Admission Test** was again field tested and revised. Forms A and B share the same General Academic (reading) component but have different Social Science and Science and Technology modules (Reading, Listening and Writing). These are presented in the Appendices

attached to the **OTESL User's Manual.** Since much of the **Post-Admission Test** field-testing required payment of subjects (for 3-hour sessions) - a very expensive item not foreseen in the project budget which had to be taken from other budget items, it was impossible to test a sufficient number of subjects with the Form B discipline-related components to carry out inter-form equating. This will have to await the gathering of further data on Form B (see proposal in Section V.)

The characteristics and intended uses of the **Placement Test** and the **Post-Admission Test** are obvious from their content, difficulty levels and format, and their target clienteles are quite different. Thus it was decided that the proposed screening procedure for students would be irrelevant. The de facto screening procedure lies in accurate descriptions of the two tests (see page 7, **OTESL User's Manual**) from which administrators will select the one appropriate for their needs.

Score reporting forms (profile score sheets and diagnostic feedback sheets) have been designed (Appendix D), to be used within and across institutions. Since the tests will be locally administered and scored, the appropriate uses of these forms will be determined by user institutions.

Statistical analysis of data continued through February, 1986, with emphasis on equating the Social Science and Science and Technology modules of Form A, on the preparation of data to be reported in the Manual on test characteristics, and on the establishment of **Post-Admission Test** cut-offs for band scores.

While the current contract with the Ministry of Education has been fulfilled, large-scale deployment of the test would ideally involve better quality production of materials (e.g. printed booklets, further editing of oral interaction training tape),and there remain several further issues of test follow-up which should be addressed by the Ministry. These are discussed in Section V.

IV: STATISTICAL CHARACTERISTICS OF OTESL

Final forms of the **Placement Test**, the Form A versions of the **Post-Admission Test** and the **Oral Interaction Test** were tested with samples of Grade 12-13, college and university ESL students, as well as with native speakers. Statistical analyses of these data are reported in the following discussion and in Appendix E, Tables 1 to 10. In general, it should be noted that the statistics are based on relatively small samples. While in the main they are quite positive, it is clear that further data should be collected as the instruments are used, and revisions should be made at a later date based on findings from larger and more diverse samples.

Means and Standard Deviations

Means and standard deviations were calculated for each OTESL component for native speakers and ESL examinees. These are reported in TABLES 1A and 1B (Appendix E).

As can be seen from the statistics for all parts of the **Post-Admission Test**, the means for native speakers are considerably higher than those for the university ESL student group, which in turn has a higher mean than the college group, with the grade 12-13 group lowest on all measures.

The ranges and standard deviations demonstrate variability within the student populations indicating that the test instruments discriminate over the proficiency range for which they were constructed.

The ESL group means for the **Placement Test** (listening and reading) are much closer to one another, reflecting the greater homogeneity of students in pre-admission ESL programmes in high schools, colleges and universities. The slight ceiling effect on these sub-tests for the samples tested is in line with feedback from ESL programmes administrators who gave the test to the effect that while the written test was sufficient in sorting lower-level students (e.g. bands 2-4), the **Placement** writing and speaking tasks were needed, as well for placing the more advanced students.

Reliability and the Standard Error of Measurement

The reliability of a testing instrument is the consistency of the results it gives. This has been calculated for the listening and reading subtests of the **Placement and Post-Admission Tests** using KR-20 estimates of internal consistency for each test component. As can be seen from Table 2 (Appendix E), reliabilities from fall 1985 - winter 1986 field-testing are .73 for the **Placement Test** and range from .85 to .90 for the **Post-Admission Test** components. The full **Post-Admission Test** versions have reliabilities of .91 and .92. Addition of the writing and speaking subtests should increase the reliability of the **Placement** measure.

The standard error of measurement is an estimate of the probable error in observed scores, assuming that because the test is not perfectly reliable an observed score may deviate somewhat from the examinee's "true score" (what an individual should achieve if the test were perfectly consistent in its measurement). If observed scores of an individual on the same test at different points in time are assumed to be normally distributed around the true score, the standard (or estimated) error of measurement, SE_m, can be used to set a confidence interval on each score.

The SE_m is important in correct test interpretation. It is given in the same units as the test score, and is used in the following way. If, for example, Student A obtains a score of 17 and the SE_m for the test is 2, student A's true score is considered (with 95% confidence) to be within the range of 13 to 21 ($\pm$ 2 SE_m above and below the obtained score). Most of the time (with 68% confidence) it is within one SE_m above or below the obtained score. Thus if 2 individuals taking the same test achieve scores within two SE_m's of each other, their scores cannot be considered to be significantly different. The SE_m's are given in Table 2 (Appendix E).

Validity

In the construction of the OTESL particular attention has been paid to face validity and content validity: that is, the careful matching of test format, content and tasks with the real-life language use situations they aim to represent, through adherence to detailed specifications. Concurrent validity data for the **Post-Admission Test** have also been analysed. Most relevant are the correlations of scores on the various instruments with teacher-provided ratings of examinee's overall language proficiency according to Carroll's 9 band General Assessment Scale (Appendix B). It must be remembered that for any given sample the teacher band ratings do not exceed a 4-5 band range; with such a limitation on possible score variance, one could not expect extremely high correlations with this measure.

In addition, TOEFL scores were available for a small number of students taking the **Post-Admission Test.** These scores represented testing sessions as much as 1 year previous to administration of the OTESL, so they can only be considered a crude indication of English proficiency at the time of OTESL administration. Furthermore, we would not expect high correlations due to the significant differences between the two instruments in testing philosophy and resulting procedures. A significant relationship would, however, be expected.

Pearson-r correlation coefficients of the teacher ratings and TOEFL scores with **Post-Admission** components scores are presented in Table 3 (Appendix E) for the student samples which were administered each pair of measures. As can be seen in Table 3, all correlations are highly significant, the lowest being that of the General Academic Test with Teacher Band ratings (.33). On the other hand, the General Academic Test correlates more highly (.49) with TOEFL scores. It may be that the speeded nature of both tests is reflected here. Teacher Band ratings have quite healthy correlations with the discipline - relevant components (.59 and .60 respectively) - certainly as high as one could reasonably expect from band ratings with their limited variance. The TOEFL scores are also significantly correlated with the discipline-relevant components, at .40 and .56 respectively. Teacher band ratings and TOEFL scores, on the other hand, have a non-significant relationship for the Social Science sample, but are correlated at .45 for the Science and Technology sample - suggesting that these measures may function differentially for science and non science students.

The above statistics plus the consistent oral feedback from teachers during field-testing indicate that the **Post-Admission Test** versions indeed rank students in the same way as do their ESL teachers. The apparent relationship but lack of identity with the TOEFL is also reassuring.

Intercorrelations among OTESL Components

Correlations among parts of the **Post-Admission Test** are presented in Tables 4 and 5, Appendix E. Since an individual's proficiency in one aspect of a language tends to overlap considerably with proficiency in another, one would expect to find substantial correlations between scores on the various components.

Table 4 presents the correlation coefficients for the General Academic Component with the Form A Social Science and Science and Technology components (.57 and .68 respectively). It also provides a coefficient for the **Oral Interaction Test** with the Form A Social Science Component (.57).

Given the nature of the General Academic component as a timed test which appears to measure fairly specific abilities versus the more comprehensive measurement of the Social Science and Science and Technology components, one would not expect to find extremely high correlations. The relative independence of the two kinds of measures is also reflected in their differential relationships to the Teacher Band Ratings and TOEFL. Thus given the reliabilities it seems obvious that both the Social Science and Science and Technology components have an amount of specific variance approximately equal to that which they have in common with the General Academic component.

The correlation between the **Oral Interaction Test** and the Social Science component is significantly different from 0 at the .05 level, but given the small sample (N=14), this relationship should be examined further when more data becomes available.

Table 5 presents the correlations among the separate parts of the **Post-Admission** versions (General Academic, reading, listening and writing). The General Academic component shows correlations of around .50 with all other components except Social Science listening (.35). Correlations of reading with listening are .51 and .50 respectively, and reading with writing .60 and .37 respectively. Correlations of each part with the total test scores are above .70 for all parts of both versions with reading making the most important contribution in both cases. These coefficients are somewhat inflated since the total includes part scores, but nonetheless indicate considerable independent contribution of each part to the total score, while showing the expected common variance.

Oral Interaction Inter-rater Reliability

The OTESL **Oral Interaction** test was field-tested continuously during its development with a total of some 80 subjects. High inter-rater agreement between the **interviewers** and **markers** was consistently found when their independent ratings were compared, with differences generally not exceeding one proficiency band. As part of the scoring procedure, consensus between the two administrators must be reached, so these independent ratings were not analysed statistically for most field-testing sessions. Table 6 (Appendix E) provides results of scoring by seven different test administrators, five of whom were trained using the OTESL Oral Interaction Test Training Videotape. (The two **interviewers** were involved in development of the instrument.) The independent ratings by the five **markers** (C-G) together with those of the **interviewers** (A and B), plus the Final Band (consensus) scores are given. As can be seen, agreement is very high.

Test Equating and Band Scores

The Social Science and Science and Technology versions of Form A were equated using the General Academic component as a common measure for the samples tested, using the procedures described in Angoff (1984). The sample sizes were 116 for Social Science and 93 for Science and Technology. The scales of equivalent scores are reported in in the **Users's Manual**, together with suggested cut-off scores for proficiency bands (Listening and Reading). It should be noted that the equivalent scores are tentative, due to the relatively small sample size and correlation between the General Academic and other components.

Test Item Characteristics

Details on item characteristics of the different components are presented in Tables 7 - 10, Appendix E. The following data are included: **item facility (IF)**, or the percentage of examinees getting the item right; **item discrimination (ID)**, an indication of the performance of higher versus lower proficiency examinees on the item; and both **biserial** and **point biserial correlations**, indices of the extent to which the ranking of examinees on each item reflects their ranking by the overall test.

It can be seen from these statistics that on the whole items fall in a reasonable difficulty range and discriminate between better and poorer students. Most of the biserial correlations are quite substantial. As part of the continuing process of revision, it was considered necessary where possible to modify either the wording of the question or the answer key for those items which appear statistically weak. Therefore these items must be evaluated in light of the results obtained in future testing.

Thus, it should be emphasized that in spite of the considerable field-testing and refinement of test materials which has taken place, more must be done. Further data collection and revision of the test instruments are urgently needed.

V. TEST DISTRIBUTION: USE, MAINTENANCE AND REVISION: A PROPOSAL

In the course of its work, the research team has become acutely aware of how easily the potential nature of testing instruments such as the OTESL **Post-Admission Test** may be undermined through breaching of security, inappropriate use, inconsistent administration and scoring procedures, and lack of revision and renewal of instruments over time. To protect the investment which has been made, and to serve, over the long term, the needs for which the battery was created, it would seem imperative to develop procedures for:

1) information dissemination and controlled test distribution to responsible institutional test users,
2) standardized training for administrators and scorers of the **Oral Interaction Test** and for scorers of the **Post-Admission Test** Writing component,
3) monitoring of test use by different institutions,
4) collection of test data for Test and Manual revision and,
5) collection of data for follow-up studies of predictive validity of the **Post-Admission Test.**

With respect to item 1), the Council of Ontario Universities or its delegate could prepare information on the test to be distributed to institutions, and could establish criteria and procedures for test distribution. While the battery should be widely available to bona fide post-secondary institutional users in Ontario, and distribution of the OTESL **Placement Test** should be straightforward, the protection of test security must be a major consideration in the case of the **OTESL Post-Admission Test.** In one possible arrangement, institutions requesting permission to use the latter test could enter into a contractual arrangement with the C.O.U. or the Ministry in which they agree to observe specific security procedures, as well as to designate

certain individuals for test administration and to make them available for training sessions (2). The institutions (and designated persons) would be responsible for collecting and forwarding the required data to the C.O.U. or its delegate (4 and 5), and for providing information, (e.g. samples of examinee performances) for periodic monitoring of test use (3). Institutions might eventually be asked to bear the cost of test materials, training sessions, and, possibly, of the centralized test maintenance system.

An existing institution, probably the Centre for Second Language Learning of the University of Ottawa, given its role to date, could be delegated the task of training (2), monitoring (3) and collection of data (4) and (5). (In the case of **Oral Interaction Test** training, the OTESL principal investigators from George Brown College and the University of Toronto would be involved.) After an adequate time lapse (one or two years) revisions would ideally be made to the instruments and to the Manual according to the data collected. This would be particularly important with respect to Version B of the **Post-Admission Test.** Eventually, as required, further forms of the test should be developed.

Initial financing of the test maintenance system should probably be in the form of a contract with the Ministry, but if ways are found to institute user fees, the operation might eventually become self-supporting.

REFERENCES

Alderson, J. Charles and A.H. Urquhart. 1985. This test is unfair: I'm not an Economist. In Hauptman, P., R. LeBlanc and M. Wesche, Second Language Performance Testing. Ottawa: University of Ottawa Press.

Angoff, W.H. 1984. Scales, Norms and Equivalent Scores. Princeton, N.J.: Educational Testing Service.

British Council, 1982. English Language Testing Service: First Report. London: The British Council.

Carroll, Brendan. 1980. Testing Communicative Performance. Oxford: Pergamon Institute of English.

Carroll, Brendan. 1981. Specifications for an English Language Service, ELT Documents: III, Issues in Language Testing. London: The British Council.

Emmett, Alan. 1985. The Associated Examining Board's Test in English for Academic Purposes. In Hauptman, P., R. Leblanc, and M. Wesche, Second Language Performance Testing. Ottawa: University of Ottawa Press.

Emmett, Alan, 1983. Test in English for Academic Purposes. Paper presented at I.U.S. The International Testing Symposium, Colchester, U.K.

Munby, John. 1978. Communicative Syllabus Design. Cambridge: Cambridge University Press.

Seaton, Ian. 1985. Issues in Validation of the English Language Testing Service (ELTS) 1976-1983. In Hauptman, R., R. Leblanc, and M. Wesche, Second Language Performance Testing. Ottawa: University of Ottawa Press.

Weir, Cyril. 1983. Identifying the Language Problems of Overseas Students in Tertiary Education in the United Kingdon. Ph.D. Dissertation, University of London, Institute of Education.

APPENDIX A

NEEDS ASSESSMENT REPORT

OTESL Needs Survey:

Before writing the test specifications, the Ontario Test of English as a Second Language (OTESL) project team undertook a needs survey of Ontario colleges and universities. There were 2 major areas of interest. First, information on admission policies for non-native speakers of English was collected along with information on the ESL programmes available to such students. To gather this information, a member of the project interviewed by telephone the admission offices of 22 colleges and 17 universities. Secondly, admission officers and ESL programme administrators were interviewed, either by telephone or in person, to gather information on the population of non-native speakers of English who were either enrolled in ESL programmes or in post-secondary programmes.

The problems with collecting the data were immediately obvious. The admission policies for non-native speakers of English at a number of institutions are not clearly defined. While visa students are commonly asked to provide proof of English proficiency, such as a TOEFL score, the policies for landed immigrants, new citizens, and francophone Canadians are often difficult to clarify. Furthermore, few institutions gather information on their population of non-native speakers of English. Thus data on language groups represented in this population and the disciplines students enter were not readily available. The information gathered was offered in most cases on a "best guess" basis by those interviewed.

A. Admission Policies

The majority of colleges and universities require a TOEFL score or a Michigan Test score for international (visa) students who have not attended an Ontario secondary school. The TOEFL is the preferred examination. Minimum score requirements vary though most institutions demand a 550 or higher. Ryerson Polytechnical Institute has the highest requirements with a 580 for students appplying for the technological programmes and a 600 for those who plan to enter communication courses. Two institutions, York University and Ryerson Polytechnical Institute, have in-house tests which students living in the Toronto area are required to take. For those institutions which accept the Michigan Test a score of 85 - 90 is required.

Policies for the admission of landed immigrants and new Canadian citizens, and visa students from Ontario secondary schools tend to be complicated. Generally, students in these categories are required to submit proof of completion of grade 13 or two to three years of high school, along with passing scores in English. Some universities require a minimum residence period in Ontario. Students who do not meet these requirements theoretically may be asked to submit a TOEFL score, but in fact this rarely happens. Landed immigrants and new citizens who are applying for admission into university programmes are generally asked to meet the admission requirements of the institution (minimum grade average and completion of pre-requisite courses) but are not asked to provide proof of English proficiency.

With the exception of University of Western Ontario, no institution asks francophone Canadians to prove English proficiency beyond providing evidence of successful completion of high school English courses.

A total of 13 institutions, a representative sample of colleges and universities, were interviewed at length on the issues summarized below. The purpose of this interview was to gather general information about the admission requirements for non-native speakers of English and about the programmes available to this group so that the OTESL project team could develop a test with the needs of administrators and of students in mind.

B. ESL Academic Preparation Courses

Administrators in seven university and college ESL academic preparation programmes were interviewed concerning the student populations of these programmes and the courses offered to these groups. According to the administrators, most of the students (80 - 90%) in these classes are visa students, primarily Arabic, Chinese and Spanish speaking. Students in these courses most commonly hope to enter programmes in business and economics, engineering, and computer science.

C. Post-Admission Courses

A number of colleges and universities offer composition courses which are required or recommended by faculties or departments. Speakers of English as a second language take a first language proficiency test which determines their need for the composition course. It is often the case that this composition course is the only post-admission course available to non-native speakers of English who have been admitted to credit programmes. In a number of institutions, however, a section of the course is designated for non-native speakers.

The population of these classes is similar to that in the academic preparation classes.

D. Testing Needs

Both admissions officers and ESL programme administrators expressed a need for a test which would determine when non-native speakers of English were qualified to enter full-time university programmes. ESL administrators are particularly interested in a test which will diagnose specific weaknesses in language proficiency. Administrators are concerned that the test not be too expensive to administer and that it accurately predict a student's academic readiness.

APPENDIX B

OTESL **POST-ADMISSION TEST** SPECIFICATIONS

Specifications for the content and item format of the OTESL **Post-Admission Test** based on a survey of academic needs of ESL students.

OTESL POST ADMISSION TEST SPECIFICATIONS

General Academic

Discourse Type	Topic	Micro-Skills/Functions +	Item Types	Skills ++
I Reading Student newspaper		20* understanding explicit info. 37 identifying the main point 39 distinguishing main point from supporting detail 41 extracting relevant information 45 skimming 46 scanning 51 transcoding information from diagrams	short answer checklist table/cells	a) Rapid reading to determine relevance of text c) Search reading for specific information d) Reading for main information

Social Sciences or Science and Technology Component

Discourse Type	Topic	Micro-Skills/Functions +	Item Types	Skills ++
IV Reading course outline textbook selection(s) case study charts and tables (source book) reference materials (table of contents, bibliography)		19* deducing lexical meaning 20* understanding explicit information 22 making inferences 37 identifying the main point 39 distinguishing main point from supporting detail 41 extracting relevant info. 44* understanding reference info. 45 skimming 46 scanning 51 transcoding information from diagrams	flowchart multiple choice true/false/ can't tell table cells	a) Rapid reading to determine relevance of text b) Reading complicated subject matter c) Search reading for specific information d) Reading for main information e) Careful reading for all information f) Critical reading to determine authors' position (SS and human.) Frequent reading formats: textbooks, duplicated notes, written questions, (class, homework, examinations), laboratory worksheets (sci. and eng.) g) Making notes from textbooks
V Listening (Reading) lecture instructions (dictation) group tutorial		20 understanding explicit info. 22 making inferences 25* expressing conceptual meaning 26 understanding sentence functions 39* distinguishing main idea from secondary info. 40 summarizing 52 transcoding info. into charts evaluation compare contrast attitudes	graphs tables cells short answer T/F/CT lists	h) Recognizing relative importance of what is being said i) Adequate note-taking from lectures (rapidity, selection, organization) j) Understanding the subject matter of the talk k) Connecting what speaker is saying to other knowledge source l) Understanding spoken description, narrative m) Understanding spoken instructions n) Following a group discussion o) Understanding abbreviations

Discipline-Relevant Component. (Continued)

Discourse Type	Topic	Micro-Skills/Functions +	Item Types	Skills ++
VI Writing essay		29 expressing sentential relations 33 developing grammatical cohesion 34 interpreting text using external information 36 using discourse markers 38 indicating the main point 39 distinguishing main ideas from secondary info. 40 summarizing 47 initiating discourse 48 maintaining discourse 49 terminating discourse 50* organizing discourse 52* transcoding information 54 relaying information		p) Range, appropriateness and correctness in written production of: - vocabulary - grammatical structures - sentence mechanics q) Organization, development of text r) Clarity of expression Frequent formats: (coursework, examinations): range of lengths from short-answer (esp. ss) to more than a paragraph (esp. eng. and sci.)

+ These skills and functions are adapted from Munby's (1978) taxonomy of micro-skills and functions.

++ Based on the extensive 3-year survey of ESL students in British universities carried out by the Associated Examining Board (Emmett 1983 1985.)

* Stars indicate skills that were particularly important for students in business studies and engineering, according to the ELTS needs analysis (Carroll 1981).

APPENDIX C

GENERAL ASSESSMENT SCALE

(Source: Carroll, Brendan J. Testing Communicative Performance, Pergamon Press, 1980, p. 134)

BAND	
9	Expert user. Communicates with authority, accuracy and style. Completely at home in idiomatic and specialist English.
8	Very good user. Presentation of subject clear and logical with fair style and appreciation of attitudinal markers. Often approaching bi-lingual competence.
7	Good user. Would cope in most situations in an English-speaking environment. Occasional slips and restrictions of language will not impede communication.
6	Competent user. Although coping well with most situations he is likely to meet, is somewhat deficient in fluency and accuracy and will have occasional misunderstandings or significant errors.
5	Modest user. Although he manages in general to communicate, often uses inaccurate or inappropriate language.
4	Marginal user. Lacking in style, fluency and accuracy, is not easy to communicate with, accent and usage cause misunderstandings. Generally can get by without serious breakdowns.
3	Extremely limited user. Does not have a working knowledge of the language for day-to-day purposes, but better than an absolute beginner. Neither productive or receptive skills allow continuous communication.
2	Intermittent user. Performance well below level of a working day-to-day knowledge of the language. Communication occurs only sporadically.
1/0	Non-user. May not even recognize with certainty which language is being used.

OTESL ORAL INTERACTION ASSESSMENT BANDS

NAME OF CANDIDATE: _____ I.D. No. _____

TESTING LOCATION: _____ DATE: _____

ASSESSMENT BANDS

above 7	Approaches bilingual competence, or better. Requires no further training in spoken English.
7	N.B. PRESENTS ACADEMIC OR COMPLEX IDEAS CLEARLY AND WITHOUT HELP. Interacts effectively and easily. Can expand and develop the point being discussed fluently and without prodding. There are hesitations and language errors but they do not interfere with communication. Speakers in this band still obviously non-native speakers of English.
6	N.B. EXPRESSES ACADEMIC OR COMPLEX IDEAS COMPREHENSIBLY AND WITHOUT MAJOR DIFFICULTY DESPITE SOME STUMBLING OR HESITATION. Is able to maintain theme of academic or complex discussion with some fluency, needs little prodding, and can develop ideas. Hesitations, errors and inappropriate language are noticeable, but do not significantly impede communication.
5	N.B. EXPRESSES SIMPLE IDEAS WELL. HAS DIFFICULTY COPING WITH ACADEMIC OR COMPLEX DISCUSSIONS WITHOUT A GREAT DEAL OF HELP. Beyond a simple discussion, may need to ask for repetition or clarification and similarly be asked for them. Requires some effort by the examiner to maintain relevance or flow of academic discussion.
4	N.B. CANNOT COPE WITH ACADEMIC OR COMPLEX DISCUSSIONS. EXPRESSES SIMPLE IDEAS COMPREHENSIBLY. Can maintain a simple discussion with some help, but hesitations, errors or inappropriate language are noticeable. Has difficulty following English at normal speed. Major problems with grammar or pronunciation.
3	N.B. HAS DIFFICULTY COPING WITH SIMPLE DISCUSSIONS. CANNOT EXPRESS SIMPLE IDEAS WITHOUT HELP. All dialogue is a drawn out affair punctuated with hesitations and severe misunderstandings. Examinee only comprehends fragments of normal speech, and has difficulties communicating ideas.
2	N.B. CANNOT COPE WITH SIMPLE DISCUSSIONS. Speaks only with words or stock phrases. Little or no ability with normal speech.
1	N.B. CANNOT COMMUNICATE IN ENGLISH AT ALL. Student is totally lost.

Examiner's Name: _____

APPENDIX D

DIAGNOSTIC FEEDBACK SHEETS

OTESL POST-ADMISSION TEST SCORE REPORT

Candidate

_____

Last Name First Name Initial

Test Date Institution

_____ _____

Month/Year

Form Version

A ___ Science & Technology ___

B ___ Social Science ___

Raw Scores

General Acad.	Reading 1	Listening	Reading 2	Writing	Total

*Proficiency Bands

Reading and Listening	Writing	Oral Interaction

* See the other side of this sheet for a description of the Proficiency Bands.

BAND	
9	Expert user. Communicates with authority, accuracy and style. Completely at home in idiomatic and specialist English.
8	Very good user. Presentation of subject clear and logical with fair style and appreciation of attitudinal markers. Often approaching bi-lingual competence.
7	Good user. Would cope in most situations in an English-speaking environment. Occasional slips and restrictions of language will not impede communication.
6	Competent user. Although coping well with most situations he is likely to meet, is somewhat deficient in fluency and accuracy and will have occasional misunderstandings or significant errors.
5	Modest user. Although he manages in general to communicate, often uses inaccurate or inappropriate language.
4	Marginal user. Lacking in style, fluency and accuracy, is not easy to communicate with, accent and usage cause misunderstandings. Generally can get by without serious breakdowns.
3	Extremely limited user. Does not have a working knowledge of the language for day-to-day purposes, but better than an absolute beginner. Neither productive or receptive skills allow continuous communication.
2	Intermittent user. Performance well below level of a working day-to-day knowledge of the language. Communication occurs only sporadically.
1/0	Non-user. May not even recognize with certainty which language is being used.

(B.J. Carroll, 1980)

OTESL WRITING EVALUATION REPORT FORM

NAME OF CANDIDATE: _____ I.D. No. _____

TESTING LOCATION: _____ DATE: _____

TEST COMPONENT _____ Placement

_____ Social Science, Form ___

_____ Science and Technology, Form ___

SCORES:

Content _____

Organization _____

Language Use _____

TOTAL _____

BAND SCORE _____

See COMBINED WRITING SCALE on other side for interpretation

COMBINED WRITING SCALE

CONTENT

7	EXCELLENT:	Knowledgeable, substantive, thorough development of thesis.
6	VERY GOOD:	Relevant to assigned topic. Accurate.
5	GOOD:	Systematic development of topic with well-structured main and subordinate themes and relevant supporting statements.
4	AVERAGE:	Mostly relevant to topic but lacks detail and supporting statements.
3	FAIR:	May lack interest but basic message gets through. Limited knowledge of subject.
2	POOR:	Backtracks and may repeat. Very little knowledge of subject.
1	VERY POOR:	A string of sentences rather than an essay. Not enough to evaluate, OR copied from text.

ORGANIZATION

7	EXCELLENT:	Fluent expression, ideas clearly stated and supported, concise.
6	VERY GOOD:	Clear and logical, cohesive
5	GOOD:	Loosely organized but main ideas stand out.
4	AVERAGE:	Logical but incomplete sequencing, choppy, redundant, not cohesive.
3	FAIR:	Structure of presentation lacks clarity. Non-fluent.
2	POOR:	Lacks logical structure and use of discourse markers.
1	VERY POOR:	No organization. Not enough to evaluate, OR copied.

LANGUAGE USE

7	EXCELLENT	Appropriate and concise English. Mastery of mechanics Sophisticated range of vocabulary.
6	VERY GOOD	Effective complex constructions. Few errors of grammar, punctuation. Effective word/idiom choice and use.
5	GOOD	Minor problems in complex sentences. Generally accurate and appropriate language, mechanics and style.
4	AVERAGE	Quite a few errors of use, mechanics and vocabulary but meaning is not confused or obscured.
3	FAIR	More frequent errors. Meaning occasionally confused or obscured, OR not an essay type format.
2	POOR	Frequent errors. Meaning often confused or obscured.
1	VERY POOR	Dominated by errors OR not enough to evaluate OR copied.

Scoring

Level 7	19 - 21 points	Level 4	10 - 12 points
Level 6	16 - 18 points	Level 3	7 - 9 points
Level 5	13 - 15 points	Level 2	4 - 6 points
		Level 1	3 points

These criteria combine B.J. Carroll's General Assessment Scale with Holly Jacob's "ESL Composition Profile," and were developed by A. Gray and L. Lebrande for OTESL.

OTESL ORAL INTERACTION DIAGNOSTIC COMMENTS

Name of candidate _____ I.D. No. _____

Testing location _____ Date _____

	Comments by examiners- TASKS 1 and 2
I. LISTENING	
- Ability to understand	
- Ability to cope with listening problems	
II. FLUENCY AND PRONUNCIATION	
- Speed/hesitancy of speech	
- Comprehensibility of pronunciation	
- Examples	
III. INTERACTIVE SKILL	
- Initiatives in keeping conversation going	
- Ability to cope with speaking problems	
- Appropriate register for task, situation	
IV. CONTENT	
- Initiation and development of topic	
- Richness and precision of content	
V. GRAMMAR AND VOCABULARY	
- Richness and precision of grammar and vocabulary	
- Accuracy of grammar and vocabulary	
- Examples	

I. LISTENING

- Ability to understand
- Ability to cope with listening problems

II. FLUENCY AND PRONUNCIATION

- Speed/hesitancy of speech
- Comprehensibility of pronunciation
- Examples

III. INTERACTIVE SKILL

- Initiatives in keeping conversation going
- Ability to cope with speaking problems
- Appropriate register for task, situation

IV. CONTENT

- Initiation and development of topic
- Richness and precision of content

V. GRAMMAR AND VOCABULARY

- Richness and precision of grammar and vocabulary
- Accuracy of grammar and vocabulary
- Examples

Comments by examiners- TASKS 3 and 4

Examiner's Name: _____

APPENDIX E
STATISTICAL TABLES

TABLE 1A: MEANS AND STANDARD DEVIATIONS FOR OTESL SCORES: NATIVE ENGLISH SPEAKERS

OTESL COMPONENT	Maximum Score	Native English Speakers				
		N	Mean	SD	Median	Range
Post-Admission Test General Academic	29*	60	26.5	2.2	27	9
Form A: Social Science (L & R)	65	31	58.1	5.0	59	22
Form A: Science & Technology (L & R)	60	30	52.8	5.0	54	21
Form A: Social Science (G.A., L & R)	94*	30	84.3	6.7	86	31
Form A: Science & Technology (G.A., L & R)	89*	30	79.5	5.72	81	26

Table 1B: MEANS AND STANDARD DEVIATIONS FOR OTESL SCORES: ESL STUDENTS

OTESL COMPONENT	Max. Score	Grade 12-13					College					University				
		N	Mean	SD	Median	Range	N	Mean	SD	Median	R	N	Mean	SD	Median	R
Placement Test (L & R)	41	47	29.9	5.3	30	21	34	28.9	4.8	30.5	16	24	31.4	3.6	31.0	14
Post-Admission Test General Academic	29*	13	16.2	4.0	16.3	16	28	17.6	5.2	17.8	25	95	20.3	4.5	20.4	19
Form A: Social Science ** (L & R)	65	18	36.4	9.0	35.5	37	--	--	--	--	--	98	44.5	10.4	45.3	53
Form A: ** Science & Technology (L & R)	60	--	--	--	--	--	33	32.5	7.4	32.0	30	42	40.5	9.2	40.5	33
Form A: Social Science ** (G.A., L & R)	94*	13	51.3	11.4	53.0	36	--	--	--	--	--	95	65.0	13.4	65.6	66
Form A: ** Science & Technology (G.A., L & R)	89*	--	--	--	--	--	28	49.3	10.9	47.5	51	42	60.6	13.9	64.5	54

* These analyses were done with the items common to all versions field-tested. Please note that the final version of the General Academic test has 36 items.

** The dictation scores are not included in these scores.

TABLE 2. RELIABILITIES AND STANDARD ERRORS OF MEASUREMENT

TEST OR SUBTEST	N	Reliability	SE_m	Maximum Possible Score*
OTESL COMPONENT				
Placement Test (L & R)	105	.73	2.6	41
Post-admission Test				
General Academic	80	.85	2.5	36
Form A: Social Science (L & R)	116	.90	3.3	65
Form A: Science & Technology (L & R)	93	.88	3.4	60
Form A: Social Science (G.A., L & R)	108	.92	4.0	94
Form A: Science & Technology (G.A., L & R)	83	.91	4.1	89

* Discrepancies here are due to revisions during field-testing.

TABLE 3 CORRELATIONS WITH OTHER MEASURES OF LANGUAGE PROFICIENCY: TEACHER RATINGS OF PROFICIENCY BANDS AND TOEFL SCORES

TEST OR SUBTEST	Teacher Band Ratings		TOEFL	
	N	Correlation Coefficient	N	Correlation Coefficient
Post-admission Test				
General Academic	57	.33	35	.49
Form A: Social Science (L & R)	64	.53	38	.37
Form A: Science & Technology (L & R)	71	.58	34	.55
Form A: Social Science (G.A., L & R)	57	.59	35	.40
Form A: Science & Technology (G.A., L & R)	61	.60	32	.56

TABLE 4 INTERCORRELATIONS AMONG OTESL COMPONENTS

COMPONENT	N	PEARSON r
General Academic with Form A: Social Science (L, R, W)	85	.57
General Academic with Form A: Science and Technology (L, R, W)	61	.68
Oral Interaction with Form A: Social Science (L, R, W)	14	.57

TABLE 5 INTERCORRELATIONS AMONG **POST ADMISSION TEST** PARTS

PART	1	2	3	4
FORM A: Social Science (N=85)				
1. General Academic	--			
2. Reading	.53	--		
3. Listening	.35	.51	--	
4. Writing	.45	.60	.40	--
TOTAL	.74	.93	.70	.76
FORM A: Science & Technology (N=61)				
1. General Academic	--			
2. Reading	.57	--		
3. Listening	.50	.50	--	
4. Writing	.52	.37	.43	--
TOTAL	.83	.84	.77	.73

TABLE 6 OTESL ORAL INTERACTION INTER-RATER RELIABILITY

	Oral Interaction Assessment Band Scores				
Student	Interviewer	Rater			Final Band
	A	C	D		
# 1	4	4	4		4
# 2	4	4	4		4
# 3	6	6	6		6
# 4	5	5	5		5
	B	E	F	G	
# 5	5	4	4	4	4
# 6	5	5	5	6	5
# 7	6	7	7	7	7

TABLE 7: ITEM CHARACTERISTICS: OTESL PLACEMENT TEST

Number of examinees = 105

Item	Point Biserial	Biserial	IF	ID
1	0.289	0.563	0.94	0.14
2	0.235	0.386	0.88	0.14
3	0.241	0.396	0.88	0.17
4	0.012	0.022	0.94	0.00
5	0.365	0.573	0.86	0.24
6	0.412	0.523	0.60	0.52
7	0.335	0.432	0.65	0.34
8	0.336	0.540	0.87	0.24
9	0.432	0.624	0.81	0.45
10	0.295	0.400	0.75	0.28
11	0.337	0.471	0.78	0.31
12	0.254	0.378	0.83	0.21
13	0.157	0.205	0.32	0.17
14	0.323	0.428	0.72	0.28
15	0.218	0.296	0.75	0.17
16	0.249	0.371	0.83	0.24
17	0.072	0.103	0.80	0.14
18	0.293	0.379	0.66	0.48
19	0.387	0.608	0.86	0.34
20	0.233	0.333	0.80	0.28
21	0.305	0.431	0.79	0.34
22	0.348	0.455	0.68	0.45
23	0.309	0.427	0.77	0.45
24	0.171	0.234	0.76	0.31
25	0.161	0.217	0.74	0.28
26	0.414	0.638	0.85	0.34
27	0.322	0.413	0.63	0.41
28	0.277	0.363	0.32	0.31
29	0.329	0.412	0.49	0.41
30	0.306	0.406	0.30	0.34
31	0.232	0.388	0.90	0.14
32	0.118	0.148	0.49	0.14
33	0.113	0.142	0.54	0.14
34	0.373	0.827	0.96	0.14
35	0.311	0.390	0.52	0.34
36	0.278	0.368	0.70	0.28
37	0.486	0.713	0.82	0.41
38	0.468	0.626	0.73	0.52
39	0.488	0.841	0.91	0.28
40	0.299	0.414	0.77	0.24
41	0.489	0.644	0.69	0.55

TABLE 8: ITEM CHARACTERISTICS: GENERAL ACADEMIC COMPONENT

Number of examinees = 80

Item	Point Biserial	Biserial	IF	ID
1	- 0.034	- 0.068	0.94	0.00
2	0.161	0.229	0.79	0.18
3	0.168	0.216	0.66	0.09
4	0.220	0.276	0.52	0.32
5	0.191	0.321	0.89	0.14
6	0.004	0.005	0.84	0.00
7	0.451	0.683	0.84	0.36
8	0.179	0.272	0.84	0.23
9	0.190	0.271	0.80	0.27
10	0.436	0.597	0.76	0.45
11	0.394	0.604	0.85	0.32
12	0.344	0.522	0.84	0.27
13	0.445	0.675	0.84	0.36
14	0.285	0.392	0.77	0.32
15	0.545	0.698	0.63	0.64
16	0.674	0.656	0.61	0.84
17	0.638	0.809	0.60	0.82
18	0.578	0.788	0.75	0.59
19	0.601	0.753	0.49	0.77
20	0.488	0.619	0.40	0.64
21	0.523	0.664	0.39	0.64
22	0.602	0.860	0.80	0.59
23	0.187	0.235	0.56	0.27
24	0.570	0.716	0.47	0.64
25	0.358	0.489	0.24	0.32
26	0.492	0.616	0.52	0.55
27	0.254	0.321	0.42	0.36
28	0.289	0.363	0.44	0.32
29	0.364	0.468	0.65	0.41
30	0.438	0.551	0.57	0.50
31	0.527	0.663	0.56	0.64
32	0.308	0.404	0.31	0.36
33	0.472	0.609	0.34	0.50
34	0.604	0.767	0.40	0.73
35	0.643	0.813	0.42	0.77
36	0.362	0.514	0.21	0.36

TABLE 9: ITEM CHARACTERISTICS: FORM A, SOCIAL SCIENCE

Number of examinees = 116

Item	Point Biserial	Biserial	IF	ID
1	0.243	0.306	0.43	0.28
2	0.079	0.136	0.90	0.06
3	0.211	0.365	0.90	0.16
4	0.163	0.265	0.88	0.13
5	0.181	0.227	0.50	0.26
6	0.322	0.411	0.63	0.38
7	0.423	0.535	0.58	0.47
8	0.357	0.450	0.44	0.38
9	0.354	0.498	0.78	0.47
10	0.227	0.392	0.91	0.16
11	0.329	0.448	0.75	0.34
12	0.121	0.151	0.53	0.22
13	0.080	0.112	0.79	0.16
14	0.154	0.204	0.70	0.16
15	0.404	0.678	0.89	0.25
16	0.428	0.537	0.46	0.56
17	0.181	0.314	0.90	0.13
18	0.444	0.566	0.62	0.50
19	0.402	0.565	0.21	0.41
20	0.342	0.434	0.61	0.41
21	0.335	0.471	0.21	0.38
22	0.337	0.512	0.85	0.25
23	0.355	0.449	0.42	0.40
24	0.550	0.699	0.39	0.59
25	0.508	0.680	0.73	0.50
26	0.456	0.573	0.55	0.63
27	0.470	0.590	0.55	0.66
28	0.344	0.454	0.70	0.38
29	0.373	0.525	0.79	0.38
30	0.297	0.383	0.65	0.31
31	0.352	0.442	0.53	0.41
32	0.330	0.415	0.47	0.41
33	0.301	0.378	0.46	0.41
34	0.346	0.476	0.76	0.31
35	0.416	0.557	0.72	0.44
36	0.470	0.592	0.43	0.59
37	0.540	0.713	0.70	0.66
38	0.503	0.674	0.72	0.63
39	0.530	0.673	0.61	0.63
40	0.575	0.769	0.73	0.66
41	0.576	0.731	0.61	0.69
42	0.594	0.790	0.71	0.66
43	0.527	0.678	0.64	0.66
44	0.566	0.711	0.54	0.66
45	0.491	0.621	0.58	0.63
46	0.498	0.632	0.61	0.63

TABLE 9: ITEM CHARACTERISTICS: FORM A, SOCIAL SCIENCE

Number of examinees = 119

Item	Point Biserial	Biserial	IF	ID
47	0.508	0.645	0.61	0.63
48	0.480	0.620	0.65	0.56
49	0.388	0.616	0.87	0.34
50	0.256	0.602	0.96	0.09
51	0.341	0.589	0.91	0.16
52	0.423	0.620	0.82	0.34
53	0.344	0.620	0.92	0.16
54	0.343	0.482	0.79	0.31
55	0.440	0.739	0.89	0.28
56	0.428	0.573	0.73	0.44
57	0.233	0.353	0.85	0.19
58	0.174	0.275	0.87	0.09
59	0.199	0.333	0.89	0.06
60	0.307	0.385	0.52	0.41
61	0.398	0.553	0.78	0.38
62	0.336	0.427	0.39	0.41
63	0.382	0.501	0.69	0.38
64	0.360	0.455	0.58	0.31
65	0.396	0.497	0.49	0.50

TABLE 10: ITEM CHARACTERISTICS: FORM A, SCIENCE AND TECHNOLOGY

Number of examinees = 119

Item	Point Biserial	Biserial	IF	ID
1	- 0.077	-0.096	0.48	-0.12
2	0.506	0.636	0.53	0.58
3	0.440	0.580	0.70	0.38
4	0.345	0.453	0.69	0.38
5	0.041	0.051	0.43	0.04
6	0.278	0.365	0.69	0.27
7	0.380	0.626	0.89	0.27
8	0.482	0.605	0.52	0.62
9	0.411	0.542	0.70	0.38
10	0.336	0.422	0.51	0.42
11	0.336	0.422	0.51	0.42
12	0.266	0.350	0.69	0.31
13	0.410	0.521	0.39	0.54
14	0.410	0.519	0.40	0.54
15	0.181	0.239	0.30	0.19
16	0.309	0.389	0.42	0.38
17	0.232	0.338	0.81	0.19
18	0.372	0.489	0.31	0.38
19	0.150	0.208	0.78	0.08
20	0.398	0.507	0.63	0.50
21	0.461	0.581	0.44	0.60
22	0.260	0.351	0.74	0.23
23	0.636	0.800	0.55	0.81
24	0.575	0.724	0.57	0.73
25	0.636	0.798	0.49	0.88
26	0.595	0.750	0.57	0.81
27	0.620	0.778	0.53	0.85
28	0.656	0.822	0.51	0.92
29	0.436	0.551	0.59	0.62
30	0.554	0.695	0.46	0.73
31	0.296	0.382	0.65	0.38
32	0.376	0.478	0.62	0.46
33	0.473	0.600	0.40	0.46
34	0.580	0.727	0.51	0.69
35	0.345	0.469	0.75	0.35
36	0.465	0.591	0.61	0.58
37	0.424	0.532	0.53	0.50
38	0.154	0.222	0.80	0.12
39	0.382	0.480	0.54	0.46
40	- 0.003	- 0.004	0.83	0.00
41	0.360	0.499	0.78	0.27
42	0.314	0.497	0.87	0.27
43	0.378	0.501	0.71	0.38
44	0.217	0.294	0.75	0.23
45	0.475	0.641	0.74	0.54
46	0.200	0.392	0.68	0.38
47	0.180	0.393	0.95	0.12
48	0.243	0.187	0.67	0.15
49	0.245	0.307	0.53	0.23
50	0.109	0.158	0.81	0.12
51	0.127	0.225	0.92	0.08
52	0.297	0.377	0.61	0.35
53	0.233	0.294	0.59	0.23
54	0.228	0.304	0.72	0.31
55	0.266	0.350	0.69	0.31
56	0.350	0.466	0.28	0.38
57	0.599	0.862	0.20	0.58
58	0.340	0.449	0.30	0.38
59	0.298	0.402	0.74	0.31
60	0.253	0.339	0.73	0.27